Kurzweil Applied Intelligence, Inc.

Developing Continuous Speech Recognition Technology that Uses Natural Language Processing Commands

During the early 1990s, tremendous market opportunities emerged for speech recognition computer technology, yet no company had been able to develop a system that could recognize natural language continuous speech commands. Development of this type of technology presented too high a level of scientific risk to attract private investment. Therefore, in 1994, Kurzweil Applied Intelligence, Inc., applied for and was awarded cost-shared funding from the Advanced Technology Program (ATP) to pursue a three-year development project. With the help of ATP funding, Kurzweil successfully developed fully operational continuous dictation technology. The technology has since been integrated into Lernout & Hauspie's VoiceXpress™ product, which allows voice control of Microsoft and Corel Office software products.

COMPOSITE PERFORMANCE SCORE

(based on a four star rating)

* * *

Research and data for Status Report 93-01-0101 were collected during July – September 2001 and April - June 2002.

Speech Recognition Technology To Provide Widespread Benefits

The benefits of developing speech recognition technology would be widespread. For example, the technology has the potential to encourage novices to use computers, and it can simplify the tasks that more experienced users encounter. Furthermore, speech recognition applications can provide expanded opportunities for the severely disabled to participate more fully in the marketplace. Finally, technology advancements offer cost savings associated with reducing repetitive motion injuries.

The technical hurdle for the industry was to craft an interface that enables personal computer (PC) users to communicate with their machines by speaking natural language commands recognized by the system. Kurzweil Applied Intelligence, Inc., proposed to create continuous speech recognition (CSR) technology referred to as Talking, Touching, and Typing ("T3"). T3 would allow a user to interact with an application by talking (saying phrases in a natural language), touching (perhaps by pen or mouse), and typing.

Existing Speech Recognition Systems Offer Limited Capabilities

Speech recognition systems are generally classified as discrete or continuous systems that are speaker dependent, independent, or adaptive. Discrete systems maintain a separate acoustic model for each word, combination of words, or phrases and are referred to as isolated (word) speech recognition (ISR). CSR systems, on the other hand, respond to a user who pronounces words, phrases, or sentences that are in a series or specific order and are dependent on each other, as if linked together.

A speaker-dependent system requires that the user record an example of the word, sentence, or phrase prior to its being recognized by the system; that is, the user "trains" the system. Some speaker-dependent systems require only that the user record a subset of system vocabulary to make the entire vocabulary recognizable. A speaker-independent system does not require any recording prior to system use. Instead, when a user identifies himself or herself, a speaker-adaptive system adapts the word, sentence, or phrase to the user's voice as the user corrects recognition errors.

ISR systems present a considerably easier task for machines than do CSR systems. Speaker-dependent systems are simpler to construct and use and are more accurate than speaker-independent systems. As a result, the focus of early voice recognition systems was primarily speaker-dependent isolated word systems that used limited vocabulary. At the time, overcoming the restrictions in the state of technology required a greater focus on human-to-computer interaction. The challenge was to identify how improved speech recognition technology could be used to support the enhancement of human interaction with machines.

The most desirable approach is where the user interacts with the PC by accomplishing a set of formal operations that are limited in scope. Kurzweil sought to accomplish this task through a natural mode of interaction that is expressive and understandable. The key was to develop a technology that balances simplicity against the robustness of the response. A command language that is too simple may not be useful. On the other hand, if the language used to give commands is too complex, it is useful only to those who understand it.

Approach Combines Human Factors Experimentation and Technology Development

Kurzweil submitted a proposal to ATP's 1993 General Competition. The company's proposed approach was to build on existing speech recognition technologies, determine the necessary parameters and restrictions required to incorporate natural language processing (NLP) commands, and integrate CSR and NLP into an interface to create the T3 technology. Some of the pivotal issues leading to this ATP award included the integration and unique combination of the technologies to be studied, as well as the understanding and commitment to human interaction as it relates to PCs in today's information-rich society. Kurzweil's proposal demonstrated an awareness of combining human factors experimentation (conducted by User Interface Engineering in a subcontractor role) and technology development throughout the project. Therefore, ATP awarded funding for a three-year period beginning in March 1994, with Kurzweil covering its indirect costs.

Kurzweil's Market Presence Enhances the Effort

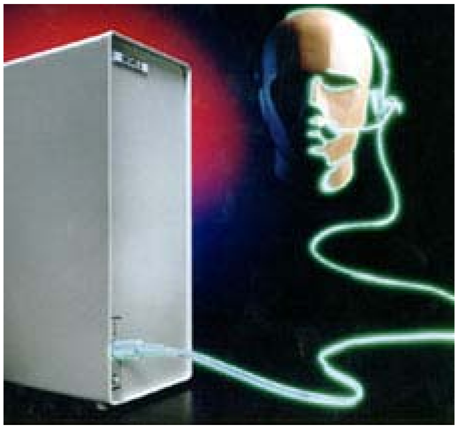
Kurzweil was founded in 1982 and proposed to use its experience, industry knowledge, and market presence to leverage the production of the interface. In 1985, the company had introduced Kurzweil Voice System, the first 1,000-word discrete-speech recognizer. This interface, adaptable to many applications, allowed the user to control the application by voice without modifying the operating system or software.

In 1987, Kurzweil introduced the first 20,000-word discrete-speech recognizer, which was incorporated into Kurzweil Voice Report software and allowed users to create structured reports by voice. A component of this technology was the Structured Report Generator (SRG). One of the key features of Kurzweil's SRG software was its ability to respond to a "trigger phrase," which is a spoken word or phrase that triggers an entire predefined report segment. Trigger phrases are designed to elicit multiple choices and alternatives as well as highlighted, fill-in-the-blank fields. The use of trigger phrases, along with word-by-word dictation, has the potential to allow users to generate custom reports by using a few spoken words. The project team built on these past efforts to develop the proposed technologies under the ATP project.

Project Succeeds in Developing Fully Functional Interface

Aided by ATP funding, the project team established a framework for technical development that included the following stages: hardware acquisition, technical planning, usability testing, and software compilation. During the technical software development stage, Kurzweil completed the construction of a continuous-speech recognizer. The company made modifications to this software throughout the project to enhance its range of accuracy. These enhancements were pivotal because the software served as the cornerstone of the spoken language interface. Kurzweil continued to make substantial progress, including the following accomplishments:

- Designed and constructed a prototype that allowed for continuous speech control of Microsoft Word and the testing of recognition accuracy



Some speaker-dependent systems require only that the user record a subset of system vocabulary to make the entire vocabulary recognizable.

- Produced both a standard and an extended speech application program interface (SAPI), in open-market format, to meet the needs of the marketplace
- Integrated several graphic user interface components with an NLP for all versions of Microsoft Word

Although the ATP project ended in February 1997, the team continued its development efforts, which resulted in a fully functional spoken language user interface system.

Acquisition and Additional Funding Advance Technology Development

In July 1997, Lernout & Hauspie acquired Kurzweil. That same year, Microsoft invested $45 million in the company, based in part on the work done in the area of a SAPI- compliant speech recognition system. The combined resources of Lernout & Hauspie and Microsoft enhanced the development and marketing of continuous command and control technologies using natural language.

> ***The team continued its development efforts, which resulted in a fully functional spoken language user interface system.***

Developing this technology involved a level of technical risk that was too high to attract private funding. Without ATP funding, Kurzweil would probably not have advanced its speech recognition technology or attracted the attention of either Lernout & Hauspie or Microsoft.

Use of Voice Technologies Continues to Grow

In September 1997, Lernout & Hauspie initially released its spoken language system as Voice Commands™, a shrink-wrapped product that allowed users to interact with PCs and word processing systems by using natural language and CSR to accomplish complex formatting and editing functions. At the time of release, however, the market was looking for a product with not only Voice Commands™ functions, but also with continuous dictation. Subsequently, the company integrated Voice Commands™ capabilities and the natural language technology into VoiceXpress™, a product that features a dictation component and grammar development tools. Lernout & Hauspie successfully marketed this product in seven languages.

These "Say it Your Way" products use patented natural language technology that enables users to dictate into a Windows-based program and allows for voice control of applications from Microsoft (Excel, PowerPoint, and Word) and Corel (WordPerfect). By 1999, Lernout & Hauspie had sold 150,000 units of VoiceXpress™ and had captured approximately 25 percent of the world market for voice recognition products.

Conclusion

In 2001, Lernout & Hauspie encountered financial troubles. The company filed for bankruptcy and was purchased for $39.5 million in assets by ScanSoft, a company known for its OmniPage optical character reader (OCR) scanning software and its digital document management software. ScanSoft intends to utilize Lernout & Hauspie's natural language technology to add dictation functions to its existing product lines. ScanSoft also is exploring putting speech recognition into automobiles and developing telephony-based products. Lernout & Hauspie was considering the development of both of these initiatives at the time of its bankruptcy.

PROJECT HIGHLIGHTS
Kurzweil Applied Intelligence, Inc.

Project Title: Developing Continuous Speech Recognition Technology that Uses Natural Language Processing Commands(Advanced Spoken Language User Interfaces for Computer Applications)

Project: To enhance human interaction with PCs by integrating advances in speech recognition using natural language technology. "Talk, Touch, and Type" interfaces allow a user to interact with an application by talking (saying phrases in a natural language), touching (perhaps by pen or mouse), and typing. The commands are not designed to handle every valid English phrase, but rather to define a simple, easily learned subset of the language to control widely used PC applications, such as word processing software and computer-aided design programs.

Duration: 3/1/1994-2/28/1997
ATP Number: 93-01-0101

Funding (in thousands):

ATP Final Cost	$1,734	72%
Participant Final Cost	664	28%
Total	$2,398	

Accomplishments: Kurzweil developed an advanced spoken language interface that is capable of responding to a user's natural language and that incorporates continuous speech recognition. The product recognizes key words in natural speech, enabling the user to request an action in multiple ways. The patented natural language technology was initially released as Voice Commands^{TM}. It was subsequently incorporated into VoiceXpress^{TM}, which the company marketed in seven languages. VoiceXpress^{TM} features natural language technology along with a dictation component and grammar development tools. Kurzweil received the following patents for technologies resulting from this ATP-funded project:

- "Speech system distinguishing dictation from commands by arbitration between continuous speech and isolated word modules"
  (No. 5,794,196: filed June 24, 1996, granted August 11, 1998)

- "System and method for remotely grouping contents of an action history stack"
  (No. 5,890,181: filed November 14, 1996, granted March 20, 1999)

- "Command parsing and rewrite system"
  (No. 6,138,098: filed June 30, 1997, granted October 24, 2000)

- "Pronoun semantic analysis system and method"
  (No. 6,125,342: filed November 18, 1997, granted November 26, 2000)

Commercialization Status: In September 1997, Lemout & Hauspie shipped its Voice Commands ^{TM} to market. Because of its restricted utility subsequently incorporated into Lernout & Hauspie's VoiceXpress^{TM}, which entered the market in 1998 and had more than 100,000 customers by 2000 (25 percent of VoiceXpress^{TM} was ATP-funded technology). In 2001, Lernout & Hauspie struggled to recover from financial troubles, a situation that led to insolvency and dissolution by the courts. ScanSoft, a company known for its OmniPage OCR scanning software and its digital document management software, acquired the Lernout & Hauspie voice recognition assets for $39.5 million in the court sale.

VoiceXpress^{TM} faced strong competition from IBM's ViaVoice products. In 2001, the IBM voice products had only a narrow sales lead over VoiceXpress^{TM}, according to NPD Intelect, but IBM benefited from Lernout & Hauspie's business failure. As of summer 2002, IBM had 53 percent of the U.S. retail sales market, compared with a 27-percent share in 2001.

Outlook: The future of Lernout & Hauspie's natural language technology appeared uncertain as the company faced dissolution. Since its purchase by ScanSoft, however, the outlook for the patented technology is more promising. ScanSoft plans to further enhance Lernout & Hauspie's technology by creating state-of-the-art digital imaging and speech and language solutions. ScanSoft also is considering putting speech recognition into automobiles and developing telephony-based products.

Composite Performance Score: * * *

Number of Employees: 100 employees at project start, 500 as of June 2002.

Company:
Scansoft, Inc.
400 5th Avenue
Waltham, MA 02451-8706

Contact: Dr. Francis Ganong
Phone: (781) 203-5110

Subcontractors:
User Interface Engineering

Research and data for Status Report 93-01-0101 were collected July – September 2001 and April – June 2002.